

Tartu Ülikool  
Loodus- ja täppisteaduste valdkond  
Matemaatika ja statistika instituut

Pille-Riin Vare

**Depressiooniravi efektiivsus ja selle võimalik geneetiline taust:  
TÜ Eesti Geenivaramu poolt läbi viidud uuringu analüüs**

Matemaatilise statistika eriala  
Bakalaureusetöö (9 EAP)

Juhendaja vanemteadur Krista Fischer

Tartu 2017

# **Depressiooniravi efektiivsus ja selle võimalik geneetiline taust: TÜ Eesti Geenivaramu poolt läbi viidud uuringu analüüs**

Bakalaureusetöö

Pille-Riin Vare

Bakalaureusetöö eesmärk on välja selgitada, kuidas muutub depressiooni raskusastet mõõtev MADRS skoor sõltuvalt ravimitest estsitalopraam ja venlafaksiin ning millised geneetilised variandid võivad mõjutada depressiooni ravimivastust. Töö alguses tuuakse välja depressiooni olemus, vajalikud geneetikaga seotud mõisted ning uuringu „Indimed – depressioon“ kirjeldus. Järgnevalt antakse ülevaade töös kasutatavatest statistilistest mudelitest: üldine lineaarne mudel (sh erijuhuna dispersioon- ja kovariatsioonanalüüsi mudelid), logistilise regressiooni mudel ning segamudel. Segamudelite puhul keskendutakse mudelitele, mis sobivad kordusmõõtmistega andmestiku korral ja kus indiviidispetsiifilised parameetrid (vabaliige, regressioonisirge tõus) on vaadeldavad juhuslike parameetritena. Viimase peatüki moodustab uuringu „Indimed – depressioon“ andmete analüüs, kus eespool tutvustatud meetodeid kasutades selgitatakse välja, millistest geneetilistest variantidest võib depressiooniravi efektiivsus 153 vähemalt mõõduka või korduva depressiooni diagnoosiga patsiendil sõltuda.

**Märksõnad:** depressioon, ühenukleotiidsed polümorfismid, andmeanalüüs

**CERCS teaduseriala:** statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika (P160)

# **The Effectiveness of the Treatment of Depression and Its Possible Genetical Background: The Analysis of Research Conducted by the Estonian Genome Centre at the University of Tartu**

Bachelor's thesis

Pille-Riin Vare

The aim of this bachelor's thesis is to find out how the change in the MADRS score, i.e. the score that shows the level of depression, is affected by the chosen treatment (escitalopram or venlafaxine), and which genetic polymorphisms may influence the depression drug response. In the introductory section the definitions of depression and important concepts of genetics are introduced, in addition to the description of the "Indimed-depression" study that is analysed in the current project. Then an overview of the statistical methods used in this paper (general linear models, including analysis of variance and covariance; logistic regression models and mixed models) is given. In the case of mixed models the author focuses on models that are used to analyse data on repeated measures, where individual-specific parameters (intercept, regression slope) can be considered as random coefficients. The last chapter analyses the data of 153 patients in the study „Indimed – depression“ using the previously described methods in order to determine which genetic polymorphisms may affect the treatment outcome of patients with at least moderate or recurring depression.

**Keywords:** depression, single nucleotide polymorphisms, data analysis

**CERCS research specialisation:** statistics, operation research, programming, actuarial mathematics (P160)

# Sisukord

<b>Sissejuhatus.....</b>	<b>5</b>
<b>1 Depressioon .....</b>	<b>6</b>
<b>2 Geneetika mõisted .....</b>	<b>8</b>
2.1 DNA .....	8
2.2 SNP.....	8
<b>3 Uuringu „IndiMed – depressioon“ eesmärk ja kirjeldus .....</b>	<b>10</b>
<b>4 Mudelid .....</b>	<b>12</b>
4.1 Dispersioonanalüüsi mudel .....	12
4.2 Kovariatsioonanalüüsi mudel.....	13
4.3 Üldine lineaarne mudel .....	14
4.4 Logistilise regressiooni mudel .....	14
4.5 Segamudel .....	15
4.6 Mudelite ja parameetrite olulisus .....	18
<b>5 Uuringu „IndiMed – depressioon“ andmete analüüs .....</b>	<b>19</b>
5.1 Ülevaade patsientidest.....	19
5.2 Tunnuste kodeeringud .....	21
5.3 Ravimid ja tausttunnused .....	22
5.4 Geneetilised variandid.....	24
5.4.1 Üldised lineaarsed mudelid .....	24
5.4.2 Kordusmõõtmiste mudelid .....	26
5.4.3 Mitmene testimine.....	28
5.5 Ebapiisav toime .....	29
<b>Kokkuvõte .....</b>	<b>31</b>
<b>Kirjanduse loetelu .....</b>	<b>33</b>

## Sissejuhatus

Depressioon on väga levinud haigus, mis võib lisaks haigestunule mõjutada ka tema lähedaste inimeste elu [1]. Ligikaudu üks viiendik inimestest kogeb elu jooksul depressiooni [1]. Depressiooniraviks leidub mitmeid erinevaid ravimeid, mida pidevalt ka täiustatakse, kuid vaatamata sellele ei saavutata alati oodatud tulemusi [2]. Järjest rohkem on kogunenud tõendeid, et geneetiline taust on ravimivastuse kujunemisel oluline [2]. Efektiivsema ravi eesmärgil on vaja täpsemalt uurida, millised geenivariandid võivad depressiooni ravimivastust mõjutada.

Bakalaureusetöö eesmärk on välja selgitada, kuidas muutub depressiooni raskusastet mõõtev MADRS skoor sõltuvalt ravimitest estsitalopraam ja venlafaksiin ning millised geneetilised variandid võivad potentsiaalselt mõjutada depressiooni ravimivastust.

Töö jaguneb kaheks osaks: teoreetiliseks ning praktiliseks. Teoreetiline osa koosneb neljast suuremast peatükist. Esimeses neist kirjeldatakse depressiooni olemust, teises tuuakse välja tähtsamad geneetikaga seotud mõisted. Ülejäänud kahes peatükis kirjeldatakse TÜ Eesti Geenivaramu poolt läbi viidud uuringut „Indimed – depressioon“, kust on pärit töös kasutatavad andmed, ning töös kasutatud mudeleid.

Praktiline pool koosneb uuringu „Indimed – depressioon“ andmete analüüsist. Kõigepealt antakse ülevaade patsientidest, seejärel kirjeldatakse üldiseid lineaarseid mudeleid, logistilise regressiooni mudeleid ning segamudeleid, mille abil püüti leida võimalikke seoseid patsientide geneetilise tausta ja ravimivastuse vahel.

Töö vormistamiseks on kasutatud tekstitöötlustarkvara Microsoft Word 2013, andmete analüüsimiseks ning jooniste tegemiseks rakendustarkvara SAS.

Autor tänab töö juhendajat Krista Fischerit kasulike ideede ning heade nõuannete eest.

# 1 Depressioon

Depressiooniks nimetatakse mitmepalgelist psüühilist häiret, mis kujuneb tavaliselt välja pikema perioodi jooksul [3]. Täpsemalt on see püsiv meeleolu alanemine, mis toob kaasa elurõõmu kadumise, energia vähenemise ning mille tõttu halveneb patsiendi elukvaliteet ning tema võime toime tulla [4].

McKenzie [5] on kirjutanud, et depressiooni korral esineb enamasti nii psüühilisi kui ka füüsilisi sümptomeid, mis isikuti varieeruvad. Psüühiliste sümptomite hulka kuuluvad tema raamatu järgi näiteks ärevus, huvi kadumine, emotsionaalne tuimus, depressiivne mõtlemine, keskendumisraskused ning enesetapumõtted. Füüsiliste sümptomitena toob McKenzie välja unega seotud probleemid, mõtlemise ja tegutsemise aeglustumise, muutused söögiisus ning väsimuse.

Järgnevad kolm lõiku tuginevad Souteri raamatule „Mis on depressioon ja kuidas sellega toime tulla“ [1]. Tänapäeva meditsiinis aktsepteeritakse asjaolu, et depressiooni ei saa päris täpselt liigitada, kuid on võimalik välja selgitada depressiooni tõsisusaste. Esiteks saab määratleda, kas tegu on primaarse haigusega või on see tekkinud mõne teise haiguse kõrvalnähtuna. Teiseks on võimalik kindlaks teha depressiooni raskusastet – kas tegu on kerge, mõõduka, raske või psühhootilise depressiooniga. Depressioonil leidub ka eriliike, näiteks unipolaarne depressioon, bipolaarne depressioon, sünnitusjärgne depressioon ning talvemasendus.

Sagedasemate depressiooni põhjustajate ning riskitegurite alla kuuluvad stressi tekitavad sündmused (lein, lahkuminek, lahutus jmt), füüsilised haigused, iseloom, soodumus suguvõsas, alkohol ning narkootikumid.

Tugevad tunded, mis depressiooni käigus tekivad, võivad hakata igapäevast elu segama. Inimene ei suuda eriti millestki rõõmu tunda ning pidevad negatiivsed mõtted võivad kaasa tuua isegi enesevigastamise või suitsiidi – seetõttu on õigesti määratud diagnoos ning ravi depressiooni puhul väga olulised.

Depressiooniraviks kasutatakse kolme liiki meetodeid – psühhofarmakoterapiat, psühhoterapiat ja füüsilisi ravimeetodeid [5]. Psühhofarmakoterapia tähendab ravimist ravimitega ja psühhoteraapia koondab enda alla erinevad rääkimisteraapia liigid [1]. Füüsiliste ravimeetodite alla kuuluvad elekterkrampravi ning valgusravi [4].

Järgnev lõik põhineb allikal [6]. Depressiooni raskusastet mõõdetakse erinevatel küsimustikel põhinevate skaalade abil. Tänapäeval on neist levinum MADRS (Montgomery–Åsberg Depression Rating Scale) skaala, mille väärtused jäävad vahemikku 0–60. Küsimustik sisaldab küsimusi kümne erineva teema kohta (näiteks pessimistlik mõtlemine, isu vähenemine, keskendumisraskused). Nende kaudu leitakse MADRS skoor, mis näitab kui tugev on depressioon. Skoor vahemikus 0–6 tähendab sümptomite puudumist, 7–19 tähistab kerget depressiooni, 20–34 näitab mõõdukat ning 35–60 tõsist depressiooni.

## 2 Geneetika mõisted

### 2.1 DNA

Rakuks nimetatakse kõige väiksemat ehituslikku ning talituslikku üksust, millel on kõik elu tunnused, ning kromosoomideks rakutuumas asuvaid geenide ehk DNA (desoksüribonukleiinhape) lõikude materiaalseid kandjaid [7]. Täpsemalt kujutab kromosoom endast DNA molekuli, mis on kaksikspiraali kujuliselt keerdunud [8]. Täiskomplekt kromosoomide, mis pärandub tervikuna ühelt vanemalt, kannab nimetust genoom [9]. Alleeliks ehk geeni esinemisvormiks nimetatakse ühte kahest või mitmest alternatiivsest geenivariandist, mis tekitab mingi tunnuse erinevaid vorme või avaldumisastmeid [8].

DNA molekul koosneb kahest ahelast, mis on vesiniksidemete abil ühendatud ning mille koostises on neli erinevat nukleotiidi (adenosiinfosfaat (A), guanosiinfosfaat (G), tsütidiinfosfaat (C) ja tümidiinfosfaat (T)) [10]. Need keerulise ehitusega ühendid on tekkinud lämmastikaluse, suhkru ja fosfaatrühma molekulide liitumisel [10]. Lämmastikaluseid on DNA ehituses samuti neli: adeniin (A), guaniin (G), tümiin (T) ja tsütosiin (C) [8]. Nukleotiidid erinevadki üksteisest ainult lämmastikaluste poolest, kuna fosfaatrühm ning suhkrujääk on kõigi nukleotiidide korral samad [10].

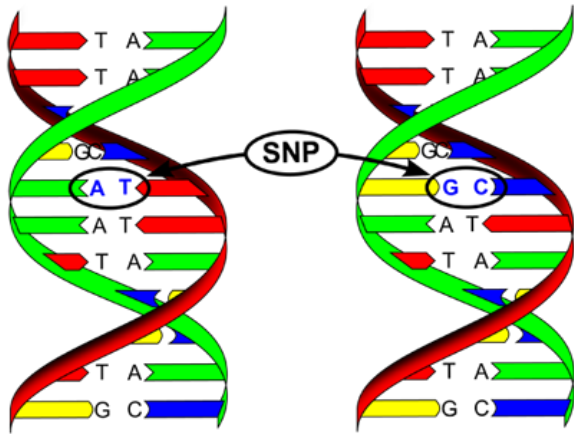
Kahe omavahel koos oleva DNA ahela aluseks on DNA komplementaarsusprintsip, mis tähendab, et nukleotiidahelad on koos kindla põhimõtte järgi: ühe ahela adenini vastas on teises ahelas alati tümiin ning guaniini vastas on tsütosiin [8].

### 2.2 SNP

Üheks põhiliseks geneetilise varieeruvuse avaldumisviisiks on inimesel SNPde olemasolu [9]. SNP ehk üksiku nukleotiidi polümorfism tähendab lihtsaimat DNA variatsiooni, mis kujutab endast ühe nukleotiidi (A, T, C või G) asendumist teise nukleotiidiga [11]. Polümorfism kujutab endast kahe või enama geneetilise variandi olemasolu populatsiooni isendite hulgas, kusjuures vähemalt kahe variandi sagedus on üle 1% [9]. Üks SNP asub keskmiselt iga 1000–2000 nukleotiidi järel [10].



Näiteks olgu DNA ahela mingis osas nukleotiidide järjestus GAATC. Kui esimese A asemel on G ehk tekib ahel GGATC, siis on tegu SNPga [10]. Eelneva näite illustreerimiseks on joonis 1, kus on kujutatud kahte DNA ahelat. Ahelate erinevus seisneb ühes nukleotiidis ning selle koha peal on tegemist üksiku nukleotiidi polümorfismiga.



Joonis 1. Näide üksiku nukleotiidi polümorfismist [12]

Järgnev lõik on refereeritud Silva Kasela bakalaureusetöö [10] põhjal. Genoomis võivad SNPd asuda nii valku kodeerivates, reguleerivates kui ka mittekodeerivates alades, kusjuures suurem osa SNPsid asub just viimastes alades. Kui üks nukleotiid on teisega asendunud geenis, mille ülesandeks on valku kodeerida, siis võib tagajärjeks olla valgu funktsiooni muutus. Üldiselt ei ole SNPd haiguste põhjustajateks, kuid nende abil on võimalik kindlaks määrata gene, millel on seos mõne haigusega.

Tuvastamaks seost geneetilise variatsiooni ning vaadeldava tunnuse või haiguse vahel, viiakse läbi ülegenoomseid assotsiatsiooniuuringuid [10]. Taoliste uuringute korral pannakse kõigepealt iga SNP jaoks paika mingi alleel efekialleeliks ning kodeeritakse SNPlle vastav tunnus kolme väärtusega: 0, 1 või 2 vastavalt efekialleelide arvule [13].

### 3 Uuringu „IndiMed – depressioon“ eesmärk ja kirjeldus

Käesolev peatükk on refereeritud Tartu Ülikooli Eesti Geenivaramu poolt koostatud uuringu protokoll [2] järgi, kui ei ole märgitud teisiti.

Antud uuringu eesmärk on välja selgitada depressiooniraviga alustanud patsientide ravitulemuse seos geneetiliste variantidega kolmes geenis (CYP2C19, CYP3A4, CYP2D6), mille ülesandeks on kodeerida antidepressante metaboliseerivaid ensüüme. Vaatluse all on patsiendid, kes on ravi alustanud antidepressantide estsitalopraami või venlafaksiiniga. Geneetiliste variantide all käsitletakse käesolevas töös SNPsid ning lisaks eespool loetletud kolmes geenis asuvatele SNPdele uuritakse veel geenides SEPSECS, CAPN2, CD247, BMP5 ja NF2 asuvaid SNPsid.

Vaadeldava uuringu puhul oli tegemist kolmekuulise jälgimisuuringuga. Jälgimisuuringu korral hinnatakse ekspositsiooni ja tulemi vahelist seost sellisel viisil, et uurija, kes on vaatleja rollis, ei määra uuritavate ekspositsioonitaset [14]. Et ekspositsiooniks on konkreetsel juhul geneetilised variandid, siis on see eeldus automaatselt täidetud. Tulemiks on depressiooni raskusastme muutus uuringu jooksul. Depressiooni raskusastet mõõdetakse antud uuringus MADRS skaala järgi.

Uuring viidi läbi Tartu Ülikooli Kliinikumi ja Põhja-Eesti Regionaalhaigla psühhiaatriaosakonna ambulatoorses vastuvõtus. Andmete kogujateks olid 36 raviarsti. Uuringusse plaaniti esialgu kaasata 300 patsienti, kuid arstidel ei õnnestunud nii suurt valimit saavutada ning uuringuga liitus lõpuks peaaegu poole vähem patsiente (153).

Uuringusse kaasamise jaoks olid patsientidele ette nähtud järgnevad kriteeriumid:

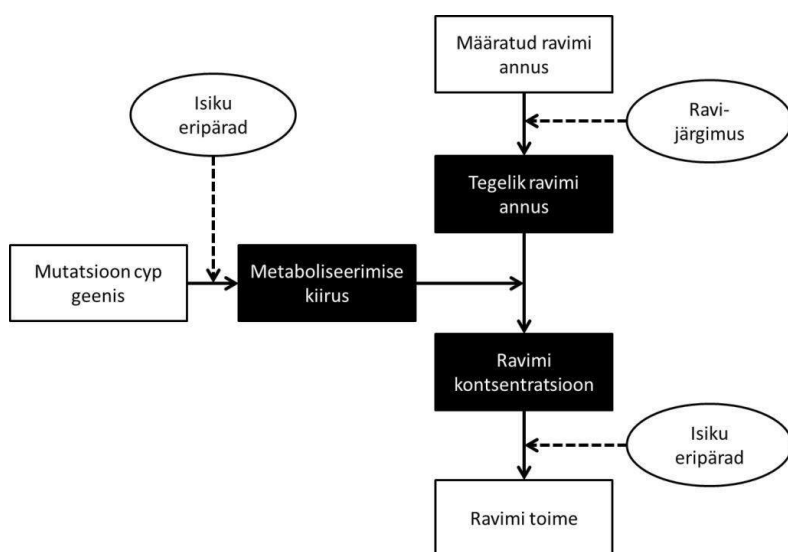
- 1) patsiendil on diagnoositud korduv depressioon või depressiooni episood;
- 2) patsiendi vanus jääb vahemikku 18–65 aastat;
- 3) patsient vajab raviarsti hinnangul antidepressiivset ravi estsitalopraami või venlafaksiiniga;
- 4) patsient ei ole vähemalt viimase kolme kuu jooksul saanud ravi antidepressiivsete ravimitega;
- 5) patsient annab uuringus osalemiseks nõusoleku.

Patsientide jaoks oli ette nähtud viis visiiti, mis toimusid kindlate ajavahemike tagant. Ajavahemike pikkust ning erinevate visiitide eesmärke on kirjeldatud tabelis 1 (lk 11).

Tabel 1. Visiitide lühikirjeldus

Visiidi number	Aeg peale esimest visiiti (päevades)	Visiidi eesmärk
1	0	Patsiendi uuringusse sobivuse hindamine, patsiendi uuringusse kaasamine, algstaatus hindamine, DNA kogumine, ravi alustamine
2	$14 \pm 2$	Ravitoime ja kõrvaltoimete hindamine, ravijärgimuse hindamine, raviskeemi stabiilsuse hindamine, vajadusel ravimi või doosi muutmine
3	$28 \pm 2$	
4	$56 \pm 2$	
5	$84 \pm 2$	

Joonisel 2 on kirjeldatud teoreetilist seost uurimise all oleva geenivariatsiooni ja antidepressandi toime vahel. Valge taustaga nelinurgad kirjeldavad neid aspekte, mille jaoks on uuringus mõeldik olemas. Musta taustaga nelinurgad esindavad selliseid tahke, mille jaoks uuringus mõeldikuid ei ole. Ovaalides olevad aspektid on sellised, mille puhul kõik toimivad komponendid ei ole teada.



Joonis 2. Teoreetiline mõjuahel geneetilise variatsiooni ja ravimivastuse vahel

Bakalaureusetöös selgitatakse välja, kuidas muutub MADRS skoor erinevaid ravimeid saavatel patsientidel ning millised geneetilised variandid võivad mõjutada depressiooni ravimivastust.

## 4 Mudelid

Käesolev peatükk on refereeritud kursuste „Andmeanalüüs II“ [15] ja „Dispersioonanalüüsi segamudelid“ konspektide [16] põhjal, kui ei ole teisiti märgitud.

### 4.1 Dispersioonanalüüsi mudel

Klassikalise dispersioonanalüüsi korral on uuritav tunnus pidev arvtunnus  $Y$  ning sõltumatud tunnused diskreetsed nn faktortunnused. Faktortunnus jaotab uuritava tunnuse rühmadeks ning faktortunnuse erinevaid väärtusi nimetatakse tasemeteks. Eelduseks on, et tasemete arv ei ole väga suur.

Dispersioonanalüüsi korral hinnatakse keskväärtuste võrdlemise abil faktori mõju uuritavale tunnusele. Olgu  $k$  üldkogumit keskväärtustega  $\mu_1, \dots, \mu_k$ .

Sellisel juhul kontrollitakse järgnevat hüpoteesipaari:

$H_0: \mu_1 = \dots = \mu_k$  ehk kõik keskväärtused on võrdsed;

$H_1: \exists i, j \mu_i \neq \mu_j (i \neq j)$ .

Üldkogumite võrdlemiseks peab olema kõigil faktorite tasemetel tehtud sõltuva tunnuse mõõtmisi. Mõõtmistulemused avalduvad mudelina:

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad (1)$$

kus  $y_{ij}$  on uuritava tunnuse väärtus  $i$ -ndas üldkogumis  $j$ -ndal objektil ( $i = 1, \dots, k$  ja  $j = 1, \dots, n_i$ ),  $\mu_i$  on  $i$ -nda üldkogumi keskväärtus ning  $\varepsilon_{ij} \sim N(0, \sigma^2)$  tähistab juhuslikku mõju.

Kirjutades  $\mu_i = \mu + \alpha_i$ , saab mudeli (1) esitada ka kujul:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

kus  $\alpha_i$  tähistab faktori  $i$ -nda taseme poolt tekitatud kõrvalekallet üldkeskmisest  $\mu$ .

Tagamaks ühesust kasutatakse erinevaid kitsendusi, näiteks  $\sum_{i=1}^k \alpha_i = 0$  või  $\alpha_1 = 0$ , kus esimesel juhul võrdub  $\mu_i$  üldkeskmisega, teisel juhul  $\mu = \mu_1$ .

Ülaltoodud hüpoteesipaari saab esitada kujul:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0;$$

$$H_1: \exists i, \alpha_i \neq 0.$$

## 4.2 Kovariatsioonanalüüsi mudel

Kovariatsioonanalüüsi mudeliks nimetatakse dispersioonanalüüsi mudelit, kuhu on lisatud pidev argument, mida nimetatakse kovariandiks. Klassikalises kovariatsioonanalüüsi mudelis ei ole faktori ja kovariandi koosmõju.

Tähistagu  $Y$  uuritavat funktsioontunnust ning olgu kovariandid  $X_1, \dots, X_s$ , ja faktorid  $A_1, \dots, A_r$ , mis tähendab, et kokku on argumente  $s + r$ . Olgu faktori  $A_i$  tasemed tähistatud vastavalt  $a_{i1}, \dots, a_{ik_i}$  ( $i = 1, \dots, r$ ), kus  $k_i$  tähistab  $i$ -nda faktori  $A_i$  tasemete arvu.

Mudeli jaoks defineeritakse indikaatorfunktsioon järgnevalt:

$$I(a_{ij}) = \begin{cases} 1, & \text{kui faktoril } A_i \text{ on tase } a_{ij}, \\ 0, & \text{teistel juhtudel.} \end{cases}$$

Mudel esitub kujul:

$$Y_i = \beta_0 + \sum_{i=1}^s \beta_i X_i + \sum_{h=1}^r \sum_{j=1}^{k_h} \gamma_{hj} I(a_{hj}) + \varepsilon_i,$$

kus  $Y_i$  on uuritava tunnuse väärtus  $i$ -ndal indiviidil,  $\beta_0$  on vabaliige, teine liidetav koosneb kovariantide väärtuste lineaarkombinatsioonidest, kolmas liidetav annab edasi faktori tasemete mõju ning  $\varepsilon \sim N(0, \sigma^2)$  tähistab mudeli jääki.

Mudeli parameetrid hinnatakse vähimruutude meetodil ning nende parameetrite arv  $p = s + \sum k_i + r + 1$ .

### 4.3 Üldine lineaarne mudel

Lineaarne mudel esitub maatrikskujul järgnevalt:

$$Y = X\beta + \varepsilon, \quad (2)$$

kus  $Y$  on  $n \times 1$  mõõtmetega uuritava tunnuse vektor,  $X$  on  $n \times p$  plaanimaatriks ( $p = k + 1$ ),  $\beta$  on  $k \times 1$  tundmatuid parameetreid sisaldav vektor ning  $\varepsilon \sim N(0, \sigma^2)$  on  $n \times 1$  mõõtmetega mudeli vigu ehk jääke sisaldav vektor.

Plaanimaatriks sisaldab pidevatele argumenttunnustele, indikaatorfunktsioonidele või koosmõjudele vastavaid veergusid.

Mudelit (2) nimetatakse üldiseks lineaarseks mudeliks, mis saab sõltuvalt argumentide tüübist olla kas regressioonanalüüsi, dispersioonanalüüsi või kovariatsioonanalüüsi mudel.

Üldjuhul on lineaarses mudelis nii pidevaid kui ka diskreetseid argumente ning neile lisaks ka erinevaid koosmõjusid. Mudel on parameetrite suhtes lineaarne ja argumentide kohta kitsendusi ei ole.

Tundmatute parameetrite jaoks kasutatakse vähimruutude meetodit, mille abil lahendatakse normaalkõrrandisüsteem  $(X^T X)\beta = X^T Y$ .

### 4.4 Logistilise regressiooni mudel

Vaatleme olukorda, kus funktsioontunnusel  $Y$  on kaks võimalikku väärtust, näiteks esineb/ei esine või on/ei ole. Sellisel juhul kodeeritakse väärtused kui 0 ja 1, kus 1 tähistab sündmuse esinemist, kusjuures  $P(Y = 1) = \pi$  ning  $P(Y = 0) = 1 - \pi$ .

Tegemist on Bernoulli jaotusega  $Y \sim B(1, \pi)$ , kus  $\pi$  on meid huvitava sündmuse tõenäosus. Keskväärus  $EY = \pi$  ning dispersioon  $DY = \pi(1 - \pi)$ . Mudel keskväärusele hindab sündmuse toimumise tõenäosust. Huvi all on uuritava tunnuse esinemise tõenäosuse  $\pi$  ja mõõdetud seletavate tunnuste vaheline seos.

Binaarse uuritava tunnuse korral kasutatakse *Logit* seosefunktsiooni

$$\omega = \text{logit}(\pi) = \ln \frac{\pi}{1-\pi}, \quad (3)$$

kus  $\frac{\pi}{1-\pi}$  tähistab sündmuse esinemise šanssi.

Logistilise regressiooni mudel, mis hindab šansi logaritmi, on defineeritud järgmiselt:

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (4)$$

kus  $\pi = P(Y = 1)$  on sündmuse esinemise tõenäosus,  $\beta_i$  ( $i = 0, \dots, k$ ) on tundmatud parameetrid ning  $x_i$  ( $i = 1, \dots, k$ ) tähistavad argumenttunnuseid.

Seosest (3) saab avaldada sündmuse esinemise tõenäosuse järgnevalt:

$$\pi = \frac{e^\omega}{1 + e^\omega}.$$

Mudeli parameetrid  $\beta_i$  ( $i = 0, \dots, k$ ) hinnatakse suurima tõepära meetodil.

## 4.5 Segamudel

Kordusmõõtmiste korral on sama tunnuse väärtust iga indiviidi korral mõõdetud korduvalt. Näiteks juhul, kui jälgitakse  $n$  patsienti seisundit, mõõtes neil kõigil  $k$  tunnust mingi jälgimisperioodi jooksul kokku  $m$  korda, siis saab rääkida kordusmõõtmistest. Üldiselt järgnevad mõõtmised ajas üksteisele. Võib öelda, et selline lähenemine on ökonoomne, kuna igalt indiviidilt saab ühe tunnuse kohta terve rea väärtusi. Tuleb arvestada seda, et ühe indiviidi tulemused on omavahel seotud.

Kordusmõõtmiste analüüsis saab eristada subjektsisest varieeruvust subjektide vahelisest varieeruvusest. Tavaliselt kasutatakse selleks segamudeleid.

Olgu esialgu tegu ühe indiviidiga, kellel uuritakse tunnust  $Y$ , mille väärtusi mõõdeti  $m$  korda ehk  $j = 1, \dots, m$ . Sellisel juhul saab kirja panna järgneva mudeli:

$$Y_j = \mu + \varepsilon_j,$$

kus  $Y_j$  on antud indiviidi uuritava tunnuse  $Y$  väärtus,  $\mu$  on antud inimese keskmine  $Y$  väärtus ning  $\varepsilon_j \sim N(0, \sigma^2)$  on  $j$ -nda mõõtmise kõrvalekalle keskmisest.

Olgu valim suurusega  $n > 1$ . Nüüd peab arvestama ka inimese järjekorranumbrit ning sellisel juhul saab esitada mudeli  $i$ -nda indiviidi jaoks järgmiselt:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad (5)$$

kus  $\mu_i$  on inimese  $i$  keskmine tunnuse  $Y$  väärtus ning  $\varepsilon_{ij} \sim N(0, \sigma^2)$  on  $i$ -nda inimese  $j$ -nda mõõtmise kõrvalekalle keskmisest ( $i = 1, \dots, n; j = 1, \dots, m_i$ , kus  $m_i$  on  $i$ -ndal indiviidil tehtud mõõtmiste arv).

Juhul, mil huvi all ei ole ühe konkreetse inimese uuritava tunnuse  $Y$  väärtus, vaid uuritava populatsiooni keskmine tunnuse  $Y$  väärtus  $\mu$ , siis avaldub iga populatsiooni kuuluva inimese keskmine  $Y$  väärtus populatsiooni keskmise kaudu järgnevalt:

$$\mu_i = \mu + \gamma_i, \quad (6)$$

kus juhuslik parameeter  $\gamma_i \sim N(0, \sigma_\gamma^2)$  on  $i$ -nda inimese omapära ehk erinevus keskmisest.

Ühendades mudelid (5) ja (6) saab mudeli kujul

$$Y_{ij} = \mu + \gamma_i + \varepsilon_{ij}. \quad (7)$$

Arvestades asjaolu, et uuritavat tunnust  $Y$  võib lisaks mõjutada ka argumenttunnus  $X$ , saab kirja panna  $i$ -nda inimese  $j$ -ndal mõõtmisel saadud uuritava tunnuse väärtuse järgmise mudeli abil:

$$Y_{ij} = \mu + \gamma_i + c_i * X_{ij} + \varepsilon_{ij}, \quad (8)$$

kus  $X_{ij}$  on  $i$ -nda inimese tunnuse  $X$  väärtus  $j$ -nda mõõtmise ajal.



Erijuhul, kui tunnuse  $X$  väärtus on iga mõõtmise korral sama, saab valemi (8) panna kirja panna järgnevalt:

$$Y_{ij} = \mu + \gamma_i + c_i * X_i + \varepsilon_{ij}, \quad (9)$$

Üldiselt võib mudel sisaldada nii selliseid argumente, mille väärtus sama indiviidi jaoks on konstantne, kui selliseid, mille väärtus võib korduvatel mõõtmistel erineda. Üheks näiteks on ajatunnus  $t_{ij}$  ( $i$ -nda indiviidi  $j$ -nda mõõtmise aeg), mille abil saab mudelisse lisada ajatunnuse.

Eeldades, et

$$c_i = c + \tau_i,$$

kus juhuslik parameeter  $\tau_i \sim N(0, \sigma_\tau^2)$ , saab järgmise mudeli:

$$Y_{ij} = \mu + c * X_{ij} + \gamma_i + \tau_i * X_{ij} + \varepsilon_{ij}.$$

Jagades juhuslikud efektid kaheks – jääkideks ( $\varepsilon_{ij}$ ) ning ülejäänuks ( $\tau_i, \gamma_i$ ), saab jääkidest ( $\varepsilon_{ij}$ ) moodustada vektori  $\varepsilon$  ning ülejäänud efektidest kombineerida ühe juhusliku valimist sõltuva vektori  $\gamma$ . Fikseeritud efektidest ( $\mu, c$ ) saab moodustada fikseeritud parameetrite vektori  $\beta$ .

Kokkuvõttes on segamudeli üldkuju järgnev:

$$Y = X\beta + Z\gamma + \varepsilon, \quad (10)$$

kus  $X$  on  $n \times p$  mõõtmetega fikseeritud efektide disainimaatriks,  $\beta$  on  $p \times 1$  fikseeritud parameetreid sisaldav vektor,  $Z$  on  $n \times z$  mõõtmetega juhuslike efektide disainimaatriks,  $\gamma$  on tundmatuid juhuslikke parameetreid sisaldav  $z \times 1$  vektor ning  $\varepsilon$  on  $n \times 1$  mõõtmetega mudeli vigu ehk jääke sisaldav vektor.

Mudeli (10) korral kehtivad veel järgnevad omadused:

- 1)  $EY = X\beta$ ,
- 2)  $E\gamma = 0$  ja  $E\varepsilon = 0$ ,
- 3)  $G := D\gamma$ ,  $R := D\varepsilon$ ,  $\gamma \perp \varepsilon$ ,
- 4)  $DY = ZGZ^T + R$ .

Segamudeli parameetreid saab hinnata üldistatud vähimruutude meetodil.

## 4.6 Mudelite ja parameetrite olulisus

Tervikuna mudeli olulisuse testimiseks kasutatakse  $F$ -statistikut ning üksikute parameetrite olulisust hinnatakse  $t$ -statistiku või  $F$ -statistiku abil.

$F$ -statistik on arvutatav suhtena, kus lugejas on mudeli poolt kirjeldatud osa funktsioontunnuse varieeruvusest ja nimetajas nn jääkvarieeruvus (see osa funktsioontunnuse varieeruvusest, mida mudel ei kirjelda). Nullhüpoteesi kehtimise korral (juhul kui funktsioontunnuse ja argumenttunnuste vahel puudub seos) on selline statistik  $F$ -jaotusega ning vabadusastmetega  $k - 1$  ja  $n - k$ , kus  $k$  on mudeli parameetrite arv ja  $n$  valimi suurus. Parameetri hinnangu ja standardvea hinnangu suhtena on leitav  $t$ -statistik ning nullhüpoteesi kehtimisel (vastav parameeter võrdub üldkogumis nulliga) on see  $t$ -jaotusega, vabadusastmete arvuga  $n - 2$  [17]. Kui  $t$ - või  $F$ -statistikule vastav olulisuse tõenäosus  $p < 0,05$ , siis loetakse vastav argument või mudel olulisusenivoo 0,05 korral statistiliselt oluliseks.

Mudelite valimiseks kasutatakse töös sammregressiooni, mille korral leitakse parim mudel samm-haaval teatud kriteeriumite alusel tunnuste hulgast argumente valides (lisades neid juurde või jättes välja). Kahanev valik (*backward*) on valikuviiis, mille korral alustatakse mudelist, mis sisaldab kõiki argumente, ning igal sammul jäetakse välja argument, millele vastava statistiku olulisuse tõenäosus ehk  $p$ -väärtus on kõige suurem. Sellist protsessi jätkatakse seni, kuni statistiliselt ei ole mudelis ühtegi mitteolulist parameetrit, kusjuures mudelist väljajäänud argumente uuesti mudelisse tagasi ei lisata.

## 5 Uuringu „IndiMed – depressioon“ andmete analüüs

Käesolevas peatükis kajastatavates analüüsides on kasutatud olulisuse nivood  $\alpha = 0,05$ .

### 5.1 Ülevaade patsientidest

Tabelist 2 (lk 20) selgub, et uuringus osales kokku 153 patsienti: 57 meest ning 96 naist. Patsientide vanus uuringuga liitudes jäi 18 ja 64 eluaasta vahele ning keskmine vanus oli 33 aastat. Kõige rohkem on patsiente, kes on 21 või 22 aastat vanad – mõlemas vanusegrupis 11 inimest. Kümme inimest on 24-aastased ning ülejäänud vanusegruppidesse jääb inimesi alla kümne.

Esimesel visiidil kirjutati patsientidele välja depressiooniravimid. Arstidel oli valida kahe antidepressandi vahel: estsitalopraam ning venlafaksiin. Estsitalopraami kirjutati välja 104 korral ja venlafaksiini määrati patsientidele vaid 49 korral (tabel 2, lk 20).

Alkoholitarbimist kirjeldav tunnus näitab, kas patsient on viimase kuu jooksul alkoholi tarbinud. Nimetatud perioodi jooksul on 114 inimest alkoholi tarbinud, mis moodustab valimist 75% (tabel 2, lk 20). Lisaks küsiti jaatavalt vastanud inimestelt, kui sageli nad keskmiselt selle perioodi jooksul alkoholiseid jooke tarbisid. Selgub, et 75 inimest tarbivad alkoholi mõnel korral kuus, 28 patsienti üks kuni kaks korda nädalas ning 11 inimest on sellised, kes tarbivad alkoholi rohkem kui kaks korda nädalas.

Uuringusse lülitamise osas pandi kirja patsientide depressiooni diagnoos, milleks sai olla vähemalt mõõdukas depressioon (F32.1, F32.2) või korduv depressioon (F33.1, F33.2). Selgub, et vähemalt mõõduka depressiooniga on 95 ning korduva depressiooniga 56 patsienti (tabel 2, lk 20). Kahe patsiendi puhul on see vastus puudulik.

Suitsetamisharjumuste kohta uuriti patsientidelt, kas nad suitsetasid uuringuga liitumise hetkel. Nendelt patsientidelt, kes suitsetasid, küsiti ka seda, kui mitmel päeval nädalas ning mitu sigaretti keskmiselt päevas suitsetatakse. Suitsetajaid on 62 ning ühe patsiendi kohta konkreetne informatsioon puudub. Tabelist 2 (lk 20) selgub, et valdav enamus suitsetab seitsmel päeval nädalas ning ühe inimese vastus on puudulik. Päevas suitsetavate sigarettide arv varieerub 1 ja 20 vahel, keskmine väärtus on 11 sigaretti. Kõige populaarsemaks vastuseks on, et päevas suitsetatakse 20 sigaretti – nii vastas 15 patsienti. 14 inimest vastas, et nad suitsetavad ühe päeva jooksul 5 sigaretti.

Tabel 2. Kaasatud patsientide andmed esimesel visiidil

Tunnus	Tunnuse väärtused	Estsitalopraam (%)	Venlafaksiin (%)
Sugu	Mees	35 (34%)	22 (45%)
	Naine	69 (66%)	27 (55%)
Vanus	Mediaan	29	29
	Keskmine (miinimum-maksimum)	34 (18-64)	33 (18-54)
Suitsetamine	Ei suitseta	65 (63%)	25 (51%)
	1-6 päeva nädalas	5 (5%)	4 (8%)
	Iga päev	32 (31%)	20 (41%)
Diagnoos	Mõõdukas depressioon	69 (66%)	26 (53%)
	Korduv depressioon	34 (33%)	22 (45%)
Alkoholitarbimine	Tarbis viimasel kuul	76 (73%)	38 (78%)
	Ei tarbinud	28 (27%)	11 (22%)

Patsientide arvu visiitidel kirjeldab tabel 3. Selgub, et uuringu katkestas 22 inimest. Kõige suurem arvuline muutus esineb kolmanda ja neljanda visiidi vahel, mil uuringu katkestas kümme patsienti. Võrreldes katkestajaid ravimigruppide kaupa saab öelda, et märkimisväärne vahe katkestajate osakaalus puudub.

Tabel 3. Patsientide arvud visiitidel

Visiidi number	Patsientide koguarv (%)	Estsitalopraam (%)	Venlafaksiin (%)
1	153 (100%)	104 (100%)	49 (100%)
2	149 (97%)	103 (99%)	46 (94%)
3	145 (95%)	100 (96%)	45 (92%)
4	135 (88%)	92 (88%)	43 (88%)
5	131 (86%)	88 (85%)	43 (88%)

Patsientidelt koguti andmeid 19 SNP kohta, mida ka analüüsid kasutatakse. Tabelis 4 (lk 21) on iga uuritava SNP korral välja toodud näiteks geeni nimetus, kus SNP asub ning täpsemad andmed kromosoomi numbri ja positsiooni kohta. Lisaks on välja toodud veel efektiivsete alleelide jaotus uuritavate patsientide korral. Osade SNPde puhul esines patsientidel mõningaid puuduvaid väärtusi.

Tabel 4. Analüüsis kasutatavad geneetilised variandid [18, 19]

						Efektialleele patsientidel		
SNP	Geeni nimetus	Kromosoom	Positsioon	Efektialleel	Alternatiivne alleel	0	1	2
rs2302566	SEPSECS	4	25124081	C	G	143	9	1
rs17599	CAPN2	1	223766378	A	CT	90	53	10
rs35864197	CD247	1	167518349	C	G	111	36	6
rs41271330	BMP5	6	55774230	A	G	1	37	115
rs79901896	NF2	22	29639032	A	G	1	24	128
rs35599367	CYP3A4	7	99768693	C	T	150	3	0
rs1065852	CYP2D6	22	42130692	C	T	10	43	98
rs1135840	CYP2D6	22	42126611	C	G	37	71	44
rs12248560	CYP2C19	10	94761900	A	CT	81	67	5
rs16947	CYP2D6	22	42127941	A	G	18	58	75
rs17885098	CYP2C19	10	94762804	C	T	2	17	134
rs28371725	CYP2D6	22	42127803	A	G	129	21	1
rs35742686	CYP2D6	22	42524243	C	CT	147	4	0
rs3892097	CYP2D6	22	42128945	A	G	100	43	8
rs4244285	CYP2C19	10	94781859	A	CG	5	29	119
rs769258	CYP2D6	22	42130761	A	G	126	25	0
rs11188072	CYP2C19	10	94759304	C	T	81	66	6
rs28633410	CYP2D6	22	42131469	C	T	133	19	0
rs5030656	CYP2D6	22	42524175	C	CCTT	146	6	0

SNPd, mis asuvad geenides CYP2D6, CYP3A4 ning CYP2C19, on varem olulisi tulemusi näidanud rahvusvahelistes uuringutes. Ülejäänud SNPd on sellised, mis on mõju näidanud Tartu Ülikooli Eesti Geenivaramu poolt läbi viidud varasemates uuringutes.

## 5.2 Tunnuste kodeeringud

Ravimi jaoks määratud tunnuse näol on tegu diskreetse kvalitatiivse tunnusega, millel on töös kasutatud andmestikus kaks väärtust: estsitalopraam ning venlafaksiin.

Sugu tähistav tunnus on andmestikus kaheväärtuseline tunnus väärtusega 1, kui tegu on mehega, ning 2, kui patsient on naissoost.

Mudelites kasutatav suitsetamist iseloomustav tunnus kirjeldab seda, kas inimene suitsetab või mitte ning selle väärtus andmestikus on vastavalt 1 või 2.

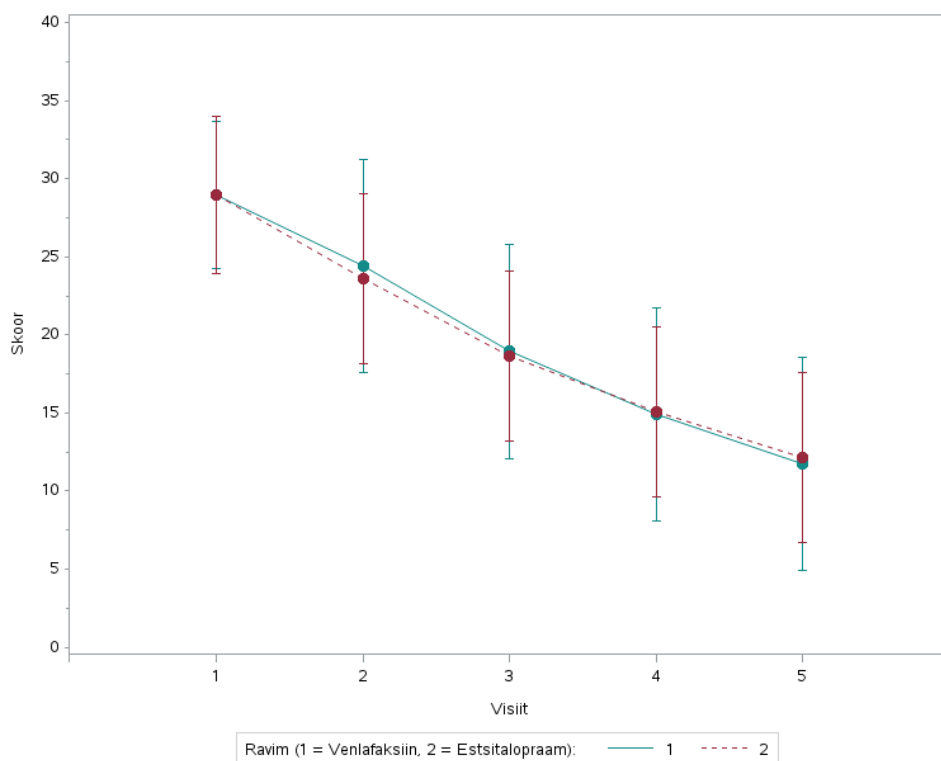
Esialgses andmestikus on depressiooni diagnoosi tähistaval tunnusel neli väärtust (1, 2, 3, 4), mis depressiooni eri vormide definitsioonide kohaselt muudeti kaheväärtuseliseks tunnuseks vastavalt 1, kui tegu on mõõduka depressiooniga, ning 2, kui tegu on korduva depressiooniga.

Alkoholitarbimist kirjeldav tunnus on samuti kaheväärtuseline ning kodeeritud väärtustega 1 ja 2 vastavalt sellele, kas patsient tarvitas või ei tarvitanud alkoholi.

Kõigile 19 uurimise all olevale SNPle vastavad tunnused on andmestikus kolme väärtusega: 0, 1 või 2 vastavalt efektilleelide arvule.

### 5.3 Ravimid ja tausttunnused

Jooniselt 3 tuleb välja, et ravimigruppide kaupa on keskmine MADRS skoor järginud sarnast langevat trendi. Selgub, et mitte ühelgi visiidil ei erine keskmine skoor ravimigruppides üle 0,8 punkti ja usalduspiirid lõikuvad kõigi visiitide korral, seega olulist erinevust MADRS skoori keskmise muutuse ja ravimi vahel ei ole. Esimesel visiidil on üldises võrdluses keskmine skoor 29 punkti, teisel visiidil 23,8 punkti, kolmandal 18,7 punkti ning neljandal ja viiendal visiidil vastavalt 15 ning 12 punkti. Seega on keskmine skoor uuringu lõpuks uuringu algusega võrreldes märgatavalt vähenenud.



Joonis 3. MADRS skoori keskmise muutus läbi visiitide 95% usalduspiiridega

MADRS skoori muutuse analüüsimiseks oli vajalik see kuidagi defineerida. Algselt plaaniti MADRS skoori muutust väljendada viienda ja esimese visiidi skoori vahena. See aga tähendaks seda, et 153 patsiendi asemel oleks 131 patsienti, kuna nii paljude andmed saadi kõigi viie visiidi kohta. Niigi väikese valimi korral on selline kadu üsna suure osakaaluga. Eeldades, et uuringu katkestanud patsientidel jäi MADRS skoor ligikaudu selliseks, nagu see oli viimasel visiidil, defineeriti skoorimuutus kui patsiendi viimase olemasoleva ja esimese visiidi vahe. See tähendab seda, et kui patsiendi kohta on olemas näiteks kolme esimese visiidi andmed, siis loetakse tema skoorimuutuseks kolmanda ja esimese visiidi MADRS skoori vahe.

Uuritakse, kas MADRS skoori muutust võivad mõjutada ka järgnevad mittegeneetilised tegurid: sugu, vanus, suitsetamine, depressiooni diagnoos, alkoholitarbimine ning ravim. Selleks koostatakse üldine lineaarne mudel kujul (2), kus uuritavaks tunnuseks on MADRS skoori muutus ning argumenttunnusteks eelnevalt loetletud mittegeneetilised tegurid. Lõpliku mudeli leidmiseks kasutatakse kahanevat valikut.

Selgub, et esimesena jääb mudelist välja vanus ( $p$ -väärtus = 0,757), teisena sugu ( $p$ -väärtus = 0,650), millele järgnevad ravim ( $p$ -väärtus = 0,409), diagnoos ( $p$ -väärtus = 0,326) ning alkoholitarbimine ( $p$ -väärtus = 0,162). Ainsana jääb mudelisse alles suitsetamist iseloomustav tunnus ( $p$ -väärtus = 0,039). Seega saab öelda, et MADRS skoori muutus sõltub suitsetamisest. Välja kirjutatuna on mudel skoorimuutusele järgmine:

$$\text{skoorimuutus} = -15,023 - 2,627 * (\text{suitsetamine} = 1), \quad (11)$$

kus suitsetamine = 1, kui inimene suitsetas uuringuga liitumise hetkel. Esialgse ja lõpliku mudeli kordajad on leitavad tabelist 5 (lk 24).

Mudeli (11) põhjal võib öelda, et suitsetaval inimesel väheneb MADRS skoor 2,627 punkti võrra rohkem kui mitesuitsetajal, sest nagu eespool mainitud, defineeriti skoorimuutus kui viimase olemasoleva ning esimese visiidi skooride vahe. Samas arvestades seda, et sellise mudelini jõudmiseks alustati 6 argumenttunnust sisaldavast mudelist, võib olla ka tegu juhuleiuga (mitmese testimise Bonferroni meetodit rakendades ei ole suitsetamisele vastav tunnus enam oluline).

Tabel 5. Skoorimuutuse mudeli parameetrite hinnangud ( $\beta$ ) koos standardvea (SE) ja  $p$ -väärtusega ( $p$ )

Parameeter	Esialgne mudel		Lõplik mudel	
	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$
Vabaliige	-16,161 (2,881)	<0,0001	-15,023 (0,804)	<0,0001
Sugu = 1	0,609 (1,350)	0,653		
Vanus	-0,017 (0,054)	0,757		
Suitsetamine = 1	-2,418 (1,314)	0,068	-2,627 (1,262)	0,039
Diagnoos = 1	-1,611 (1,407)	0,254		
Alkoholitarbimine = 1	1,969 (1,548)	0,206		
Ravim = estsitalopraam	1,225 (1,417)	0,389		

Eespool defineeritud MADRS skoori muutusele otsustati lisaks luua uus tunnus *muutus*, mis defineeriti eespool mainitud skoorimuutuse ning esimese visiidi MADRS skoori suhtena, kuna see aitab korrigeerida võimalikku sõltuvust skoori algsest väärtusest. Eelnevaga sarnaselt koostatakse üldine lineaarne mudel kujul (2), kus funktsioontunnuseks on uus tunnus *muutus* ning argumenttunnusteks sugu, vanus, suitsetamine, diagnoos, alkoholitarbimine ning ravim. Lõpliku mudeli leidmiseks kasutatakse kahanevat valikut. Sellele tunnusele ei osutu hinnatud mudelis ühegi tunnuse mõju oluliseks.

## 5.4 Geneetilised variandid

### 5.4.1 Üldised lineaarsed mudelid

Lisaks otsustati üldiste lineaarsete mudelite abil uurida ka seoseid skoorimuutuse ning erinevate SNPde vahel.

Iga SNP puhul hinnatakse mudel, kus funktsioontunnuseks on skoorimuutus ning argumenttunnuseks vastav SNP, kusjuures mudelid luuakse kõigi 19 SNP jaoks, mille kohta on andmestikus andmed olemas. Mudelid hinnatakse eraldi ravigruppide kaupa kui ka üldiselt. Tulemused on kajastatud tabelis 6 (lk 25).

Tabelist 6 (lk 25) selgub, et kogu valimi korral jäävad oluliseks kolm SNPd: rs2302566, rs79901896 ja rs17885098. Estsitalopraami tarvitajate hulgas on olulised kaks SNPd – rs2302566 ja rs50300656, millest esimene on oluline ka kogu valimi korral. Venlafaksiini korral ei osutu ükski SNP oluliseks. Geneetilised variandid rs17885098 ja



rs5030656 asuvad vastavalt geenides CYP2C19 ning CYP2D6, mille seos depressiooni ravimivastusega on leitud ka varem rahvusvahelise uuringu käigus.

Kuna eelnevast selgus, et suitsetamist kirjeldav tunnus mõjutab oluliselt skoori muutumist, siis luuakse ka mudelid, milles funktsioontunnuseks on MADRS skoori muutus ning argumenttunnusteks on SNP ja suitsetamist kirjeldav tunnus. Ka antud juhul kasutatakse tahapoolse valikut ja vaadeldakse kõigi patsientide andmeid. Selgub, et geneetilise variandi rs2302566 korral jäävad mudelis oluliseks SNP ise ( $p$ -väärtus = 0,009) kui ka suitsetamine ( $p$ -väärtus = 0,045). Geneetilise variandi rs79901896 korral jääb lõplikusse mudelisse alles olulisena SNP ( $p$ -väärtus = 0,014). Teiste SNPde korral olulisi tulemusi ei tule.

Tabel 6. Üldiste lineaarsete mudelite parameetrite hinnangud ( $\beta$ ) koos standardvea (SE) ja  $p$ -väärtusega ( $p$ )

SNP	Kogu valim		Estsitalopraam		Venlafaksiin	
	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$
rs2302566	7,016 (2,570)	<b>0,007</b>	8,986 (3,199)	<b>0,006</b>	3,140 (4,279)	0,467
rs17599	-1,513 (1,005)	0,134	-1,976 (1,179)	0,097	-0,566 (2,043)	0,783
rs35864197	-1,728 (1,144)	0,134	-1,508 (1,452)	0,302	-1,978 (1,855)	0,292
rs41271330	-0,053 (1,391)	0,970	-0,638 (1,668)	0,703	1,872 (2,562)	0,469
rs79901896	3,843 (1,549)	<b>0,014</b>	3,784 (2,016)	0,064	3,670 (2,430)	0,138
rs35599367	-7,674 (4,423)	0,085	-	-	-7,202 (4,166)	0,091
rs1065852	0,015 (1,032)	0,884	0,321 (1,212)	0,791	0,078 (2,104)	0,971
rs1135840	0,409 (0,871)	0,639	1,435 (1,091)	0,192	-1,569 (1,404)	0,270
rs12248560	1,142 (1,108)	0,202	2,459 (1,428)	0,088	-0,156 (1,733)	0,929
rs16947	-0,610 (0,927)	0,512	-2,204 (1,223)	0,075	1,524 (1,400)	0,282
rs17885098	3,262 (1,609)	<b>0,045</b>	3,111 (1,981)	0,120	3,923 (2,775)	0,165
rs28371725	1,434 (1,690)	0,398	1,105 (2,122)	0,604	2,342 (2,782)	0,404
rs35742686	-2,187 (3,904)	0,576	-2,601 (4,057)	0,523	-	-
rs3892097	0,061 (1,078)	0,955	-0,023 (1,279)	0,986	-0,078 (2,104)	0,971
rs4244285	-1,025 (1,227)	0,405	-1,772 (1,460)	0,228	1,389 (2,290)	0,547
rs769258	-1,202 (1,690)	0,478	0,246 (2,368)	0,918	-2,569 (2,390)	0,288
rs11188072	1,507 (1,085)	0,167	2,515 (1,379)	0,071	-0,156 (1,733)	0,929
rs28633410	-0,329 (1,933)	0,865	2,554 (2,630)	0,334	-3,862 (2,743)	0,166
rs5030656	-5,286 (3,162)	0,097	-7,571 (3,541)	<b>0,035</b>	4,886 (7,254)	0,504

#### 5.4.2 Kordusmõõtmiste mudelid

Patsientidelt koguti andmeid 19 erineva SNP ehk geneetilise variandi kohta. Selgitamaks välja, kas mõne SNP korral on aja ja SNP vahel oleva koosmõju ning MADRS skoori muutuse vahel oluline seos, vaadeldakse esialgu segamudeleid kujul (10) MADRS skoori muutusele. Iga kord lisatakse mudelisse argumenttunnusteks SNP, aeg ning SNP ja aja koosmõju. Sel moel tuleb välja viis geneetilist varianti, mille korral nende koosmõju ajaga on kõigi patsientide andmete põhjal oluline. Tabelist 7 (lk 27) selgub, et koosmõju aja ning SNP vahel osutub oluliseks järgnevate SNPde korral: rs2302566, rs35864197, rs79901896, rs35599367 ning rs17885098.

Lisaks kogu valimile vaadeldake MADRS skoori muutuse ja SNP vahelisi seoseid sõltuvalt ravimigrupist. Selleks tehakse mudelid kujul (10), kus argumenttunnusteks on SNP, aeg ning SNP ja aja koosmõju. Tabelist 7 (lk 27) selgub, et estsitalopraami korral jääb oluliseks seitse SNPd, millest kaks (rs2302566, rs17885098) on samad, mis jäid oluliseks ka kõigi patsientide korral. Lisaks selgub, et venlafaksiini korral osutub oluliseks ainult üks SNP (rs35599367), mis kogu valimi ega estsitalopraami tarvitajate grupi korral oluline ei jäänud.

Leitud segamudelitesse proovitakse lisada ka eespool kasutatud tausttunnuseid (sugu, vanus, suitsetamine, diagnoos, alkohol). Täpsemalt täiendatakse kogu valimi neid viit segamudelit, mille korral SNP ja aja koosmõju oluline jäi. Selgub, et mitte ühelgi korral ei jää ükski tausttunnus statistiliselt oluline.

Leidmaks, kas mõni tausttunnus segamudelil ilma SNP ja SNP ning aja koosmõjuta ka oluline võiks olla, tehakse segamudel skoorimuutusele, kus argumenttunnusteks vaid uurimise all olevad tausttunnused sugu, vanus, suitsetamine, diagnoos, alkoholitarbimine. Sel moel ei osutu samuti ükski tunnus oluliseks.

Tabel 7. Kordusmõõtmiste mudelites SNP ja aja koosmõjule vastavate parameetrite hinnangud ( $\beta$ ) koos standardvea (SE) ja  $p$ -väärtusega ( $p$ )

SNP	Kogu valim		Estsitalopraam		Venlafaksiin	
	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$
rs2302566	1,320 (0,534)	<b>0,014</b>	1,472 (0,738)	<b>0,046</b>	1,085 (0,758)	0,154
rs17599	-0,283 (0,204)	0,165	-0,3430 (0,244)	0,161	-0,248 (0,389)	0,524
rs35864197	-0,513 (0,2301)	<b>0,027</b>	-0,492 (0,299)	0,101	-0,529 (0,354)	0,137
rs41271330	-0,187 (0,278)	0,503	-0,381 (0,337)	0,259	0,442 (0,496)	0,374
rs79901896	0,747 (0,307)	<b>0,015</b>	0,773 (0,401)	0,055	0,641 (0,468)	0,173
rs35599367	-1,790 (0,855)	<b>0,037</b>	-	-	1,697 (0,798)	<b>0,035</b>
rs1065852	-0,020 (0,205)	0,921	-0,008 (0,243)	0,975	0,717 (0,408)	0,861
rs1135840	0,156 (0,175)	0,373	0,4722 (0,223)	<b>0,035</b>	-0,408 (0,272)	0,136
rs12248560	0,314 (0,221)	0,155	0,621 (0,291)	<b>0,033</b>	-0,129 (0,329)	0,694
rs16947	-0,173 (0,184)	0,346	-0,624 (0,251)	<b>0,013</b>	0,364 (0,268)	0,177
rs17885098	0,756 (0,321)	<b>0,019</b>	0,808 (0,401)	<b>0,045</b>	0,683 (0,531)	0,120
rs28371725	0,164 (0,340)	0,629	0,003 (0,430)	0,995	0,487 (0,546)	0,374
rs35742686	-0,640 (0,795)	0,422	-0,740 (0,829)	0,373	-	-
rs3892097	0,054 (0,216)	0,802	0,061 (0,258)	0,812	-0,072 (0,408)	0,861
rs4244285	-0,321 (0,241)	0,184	-0,501 (0,291)	0,086	0,229 (0,430)	0,597
rs769258	0,018 (0,326)	0,955	0,404 (0,456)	0,377	-0,352 (0,457)	0,442
rs11188072	0,350 (0,216)	0,105	0,653 (0,280)	<b>0,020</b>	-0,129 (0,329)	0,694
rs28633410	-0,141 (0,390)	0,717	0,671 (0,564)	0,235	-0,903 (0,520)	0,084
rs5030656	-1,012 (0,613)	0,099	-1,654 (0,695)	<b>0,018</b>	1,884 (1,362)	0,169

Huvi pakub ka see, kas segamudelitesse kujul (10) mitmeid tabelis 7 olulisi SNPsid korraga lisades jääb mõni neist oluliseks. Segamudeleid uuritakse nii kõigi andmete kui ka estsitalopraami ravigrupi korral eraldi. Venlafaksiini ravigrupi korral pole võimalik korraga mitme SNP mõju uurida, sest tabelist 7 on näha, et selle ravimi korral on oluline ainult üks SNP. Lõplike mudelite leidmiseks, kus oleks ainult olulised SNPd sees, kasutatakse kahanevat valikut. Funktsioontunnuseks on mõlemal juhul eespool defineeritud MADRS skoori muutus ning argumenttunnusteks on kogu valimi korral SNPd rs2302566, rs35864197, rs79901896, rs35599367, rs17885098 ning estsitalopraami ravigrupi korral SNPd rs2302566, rs1135840, rs12248560, rs16947, rs17885098, rs11188072, rs5030656.

Kõigi patsientide andmeid korraga vaadates selgub, et üks SNP jääb oluliseks, täpsemalt rs35864197. Estsitalopraami ravigruppi kuulunud patsientide korral jääb samuti ainult üks SNP oluliseks: rs16947.

### 5.4.3 Mitmene testimine

Käesolev lõik põhineb Liina Tamme bakalaureusetööl [20]. Mitmese testimise probleem seisneb selles, et kui korraga testida palju erinevaid hüpoteese, siis leitakse mõni statistiliselt oluline seos ka juhul, kui see tegelikult ei eksisteeri. Mitmese testimise meetodeid on mitmeid, millest üks levinumaid on Bonferroni meetod. Antud meetodi aluseks on Bonferroni võrratus, mille põhimõtteks on, et I liiki vea tõenäosus on kõikide testide korral väiksem või võrdne kui I liiki vea tegemise tõenäosuste summa iga testi korral eraldi. Kui on vaja, et kogu eksperimendi ulatuses oleks I liiki vea tegemise tõenäosus  $\pi$ , siis peab iga testi korral I liiki vea tegemise tõenäosus olema

$$\alpha = \frac{\pi}{n},$$

kus  $n$  tähistab testide koguarvu.

Arvestades, et kokku testitakse antud juhul 19 SNP mõju, saab otsustada nende olulisuse üle vaid siis, kui on arvestatud mitmese testimisega. Testitud SNPde hulgas on neid, mis asuvad genoomis üksteise lähedal ning on seega korreleeritud, mistõttu ei saa teste sõltumatuteks lugeda. Kasutades Bonferroni meetodit, peaks võrdlema kõiki olulisuse tõenäosusi  $0,05/19 = 0,0026$ -ga, et I liiki vea tõenäosus üle kõigi testide jääks alla 5%.

Seda kriteeriumit kasutades ei saa mitte ühegi SNP mõju depressiooni ravimivastusele lugeda oluliseks. Siiski võib tähele panna, et  $p$ -väärtuseid, mis on alla 0,05 (või ka alla 0,1), on rohkem, kui juhuslikult võiks oodata.

Hindamaks ravimivastusega seotud SNPde võimalikku arvu, eemaldati SNPd rs1135840, rs16947, rs3892097, rs11188072, rs28633410, mille korrelatsioon mõne teise SNPga ületab 0,25, ning tehti eeldus, et ülejäänud 14 SNPd on statistiliselt sõltumatud. Kui neist mitte ükski ei oleks ravimivastusega seotud, siis võiks alla 0,05 jäävate  $p$ -väärtuste arv olla binoomjaotusega juhuslik suurus  $P_5 \sim \text{Bin}(14; 0,05)$ . Sellisel juhul  $P(P_5 > 2) = 0,03$  ehk tõenäosus, et juhuslikult tekib 3 või enam 0,05-st väiksemat  $p$ -väärtust, on umbes 3%.

Estsitalopraami puhul leiti nende 14 SNP puhul neli ja kahe ravimi andmeid kokku analüüsides viis  $p$ -väärtust, mis on väiksemad kui 0,05. Seega võib eeldada, et vastavate SNPde hulgas on mõni, mis on ka tegelikult ravimivastusega seotud.

Sama tulemuseni jõuti ka permutatsioonitesti abil, mis kujutab endast teststatistiku nullhüpoteesile vastava jaotuse leidmist arvutades teststatistiku väärtused andmete kõikvõimalike ümberpaigutuste korral [21]. Juhuslikult ravimivastuste permuteerimisel tekitatud valimites leiti kolm või enam olulist  $p$ -väärtust vaid 3% valimite korral. Permutatsioonitesti ei viinud läbi käesoleva töö autor, kuna see oleks olnud bakalaureusetöö jaoks liiga mahukas.

## 5.5 Ebapiisav toime

Alates teisest visiidist on võimalik hinnata, kas patsiendil esineb ravimi osas ebapiisav toime, mille tagajärjel oli vaja näiteks ravimiannust tõsta. Uurimaks, kas ebapiisava toime esinemine sõltub ravimist, loodi andmestikku uus tunnus *ebapiisav*, mille väärtused olid 1 ja 0 vastavalt sellele, kas patsiendil esines vähemal ühel visiidil ebapiisav toime või mitte.

Uuritava küsimuse jaoks koostatakse logistilise regressiooni mudel kujul (4), mis antud juhul näeks ühe argumenttunnuse tõttu välja järgnev:

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1,$$

kus  $\pi = P(\text{ebapiisav toime} = 1)$  on ebapiisava toime esinemise tõenäosus ning  $x_1$  tähistab ravimit.

Selgub, et  $p$ -väärtus = 0,002, mis tähendab seda, et ebapiisava toime esinemine sõltub ravimist. Lõplik mudeli kuju on järgmine:

$$\ln \frac{\pi}{1-\pi} = 0,4322 - 0,5864 (\text{ravim} = \text{estsitalopraam}).$$

Osutub, et šanss ebapiisava toime esinemiseks estsitalopraami tarvitajatel on  $e^{-0,5864} = 0,556$  korda väiksem kui venlafaksiini tarvitajatel. Teiste sõnadega tähendab see seda, et venlafaksiini tarvitajate šanss ebapiisava toime esinemiseks on ligikaudu  $e^{0,5864} = 1,798$  korda suurem.

Täpsemalt uurides selgub, et ebapiisav toime esineb 84 patsiendil, kellest 48 kuulub estsitalopraami ravigruppi ning 36 venlafaksiini ravigruppi. Arvestades teadmist, et estsitalopraami tarvitajaid on kokku 104 ning venlafaksiini tarvitajaid 49, siis selgub, et 73% venlafaksiini ning 46% estsitalopraami tarvitajatel täheldati ravimi ebapiisavat toimet.

Lisaks uuritakse, kas ebapiisav toime sõltub mõnest SNPst või SNP ja ravimi koosmõjust. Küsimusele vastuse saamiseks luuakse mudeleid kujul (4), kus argumenttunnusteks lisatakse ravim, SNP ning SNP ja ravimi koosmõju. Oluliste tunnustega lõplike mudelite leidmiseks kasutatakse kahanevat valikut. Kuna eesmärgiks on leida olulisi SNPsid või SNP ja ravimi koosmõjusid, siis neid olukordi, kus lõplikkusse mudelisse jääb alles ainult ravim, eraldi kajastama ei hakata.

Tabel 8. Logistilise regressiooni mudelite parameetrite hinnangud ( $\beta$ ) koos standardvea (SE) ja  $p$ -väärtusega ( $p$ )

SNP	SNP		ravim = estsitalopraam		SNP*(ravim =estsitalopraam)	
	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$	$\beta$ (SE)	$p$
rs1065852	-	-	-	-	0,380 (0,111)	0,001
rs1135840	-	-	1,502 (0,359)	<0,0001	-0,803 (0,254)	0,002
rs16947	-	-	-	-	0,479 (0,132)	0,0003
rs35864197	0,709 (0,329)	0,031	0,6370 (0,196)	0,001	-	-

Tabelist 8, kus on kõigi patsientide andmete põhjal loodud mudelite andmed, on näha, et nelja SNP korral õnnestub leida mudeleid, kus lõplikus mudelis jääb oluliseks ka SNP või SNP ja ravimi koosmõju. SNPde rs1065852 ja rs16947 korral jääb lõplikusse mudelisse alles ainult SNP ja ravimi koosmõju. SNP ning ravim eraldi jäävad oluliseks rs35864197 korral ning ravim ning ravimi ja SNP koosmõju jäävad alles rs1135840 korral.

Seega saab tabeli 8 põhjal öelda, et efektilleelid SNPdes rs1065852 ja rs16947 suurendavad ebapiisava toime esinemise šanssi estsitalopraami tarvitajatel. SNP rs35864197 efektilleelid suurendavad ebapiisava toime esinemise šanssi mõlema ravimi korral.

Selgitamaks välja, kas SNPd ka ravigruppides eraldi ebapiisava toime esinemist võivad mõjutada, uuritakse mudeleid kujul (4), kus funktsioontunnuseks on tunnus *ebapiisav* ning argumenttunnuseks konkreetne SNP. Selgub, et venlafaksiini korral ei jää ükski SNP oluline. Estsitalopraami ravigrupis jäävad oluliseks kaks SNPd: rs35864197, rs1135840, kusjuures rs35864197 efektilleelid suurendavad ebapiisava toime esinemise šanssi ning rs1135840 efektilleelid vähendavad ebapiisava toime esinemise šanssi.

## Kokkuvõte

Bakalaureusetöö eesmärk oli kindlaks teha, kuidas muutub depressiooni raskusastet mõõtev MADRS skoor sõltuvalt ravimitest estsitalopraam ja venlafaksiin ning millistel geneetilistel variantidel võib esineda mõju depressiooni ravimivastusele. Kasutatavad andmed olid pärit Tartu Ülikooli Eesti Geenivaramu poolt läbi viidud uuringust „Indimed – depressioon“. Analüüsimiseks hinnati segamudeleid kordusmõõtmistele, logistilise regressiooni mudeleid ravimi ebapiisava toime esinemise šansile ning üldiseid lineaarseid mudeleid depressiooni raskusastme muutusele.

Töö teoreetilises osas anti ülevaade depressiooni olemusest, kirjeldati vajalikke geneetikaga seotud mõisteid, uuringut „Indimed – depressioon“ ning töös kasutatud mudeleid. Praktilises osas kirjeldati andmeid ning koostati töö eesmärkide saavutamiseks vajalikud mudeleid.

Selgus, et ravimigruppide kaupa võrreldes ei erinenud MADRS skoori keskmine oluliselt mitte ühelgi visiidil – erinevused ei ületanud 0,8 punkti. Üldises võrdluses oli keskmine skoor esimesel visiidil 29 punkti ning viimasel märgatavalt väiksem: 12 punkti.

Otsides mittegeneetiliste tegurite (sugu, vanus, suitsetamine, depressiooni diagnoos, alkohoolitarbimine, ravim) mõju MADRS skoori muutusele, selgus, et ainukesena võib mõju olla suitsetamist kirjeldav tunnusel – indiviidil, kes suitsetas, langes MADRS skoor mittersuitsetajatega võrreldes 2,6 punkti võrra rohkem.

Uurides, kas skoorimuutus võib sõltuda mõnest SNPst, selgus, et kõigi patsientide puhul oli potentsiaalseid SNPsid kolm (rs2302566, rs79901896, rs17885098), estsitalopraami korral kaks (rs2302566, rs5030656) ning venlafaksiini korral ei tulnud ükski oluline. Oluliseks osutunud geneetilised variandid rs17885098 ja rs5030656 asuvad vastavalt geenides CYP2C19 ning CYP2D6, mille seos depressiooni ravimivastusega on leitud ka varem rahvusvahelise uuringu käigus.

Kordusmõõtmiste korral tuli välja, et aja ning SNP koosmõju osutus oluliseks kõigi patsientide andmete korral viies SNPs: rs2302566, rs35864197, rs79901896, rs35599367 ja rs17885098. Estsitalopraami korral jäi oluliseks seitse SNPd, millest kaks (rs2302566, rs17885098) olid samad, mis kõigi andmete korral. Venlafaksiini korral osutus oluliseks ainult üks SNP (rs35599367).

Kuigi SNPde korral tekkis mitmese testimise probleem, siis selgus, et kui kõik oluliseks osutunud SNPd päriselt ravimivastust ei pruugi mõjutada, leidub mõni neist, mis on ka tegelikult ravimivastusega seotud.

Osutus, et ravimi ebapiisava toime esinemine sõltub ravimist. Šanss, et venlafaksiini tarvitajatel esineb ravimi ebapiisav toime, on 1,8 korda suurem kui estsitalopraami tarvitajatel. Analüüs näitas, et 73% venlafaksiini ning 46% estsitalopraami tarvitajatel esines ravimi ebapiisav toime.

Bakalaureusetöö võiks olla aluseks edasisele geneetika valdkonna teadustööle. Tulemuste lahtimõtestamiseks bioloogia ja meditsiini seisukohalt on edaspidi vaja geneetikute ja statistikute koostööd ning suuremahulisemaid uuringuid.



## Kirjanduse loetelu

- [1] Souter, K. (2015). *Mis on depressioon ja kuidas sellega toime tulla*. Tallinn: Tammeraamat.
- [2] Tartu Ülikooli Eesti Geenivaramu. (2014). Uuringu „Tavapraktikas esitsitalopraami või venlafaksiiniga depressiooni ravi alustanud patsientide ravitulemuse seos mutatsioonidega antidepressante metaboliseerivaid ensüüme kodeerivates geenides (CYP2C19, CYP3A4, CYP2D6)“ protokoll.
- [3] Huttunen, M. O. (2006). *Depressioon*. Tallinn: AS Medicina.
- [4] Tartu Ülikooli Kliinikum. (2015). *Depressioon*. Infoleht. [https://haiglateliit.ee/wp-content/uploads/2015/04/patsiendile\\_014\\_depressioon.pdf](https://haiglateliit.ee/wp-content/uploads/2015/04/patsiendile_014_depressioon.pdf) (09.05.2017)
- [5] McKenzie, K. (2010). *Mis on depressioon*. Tallinn: Eesti Ajalehed AS.
- [6] McDowell, I. (2006) *Measuring Health: A Guide to Rating Scales and Questionnaires, Third Edition*. New York: Oxford University Press. <http://www.fundacion-salto.org/documentos/Measuring%20Health.pdf> (09.05.2017).
- [7] Järvalt, H. (2009). *Bioloogia lühikursus gümnaasiumile*. Tallinn: Avita.
- [8] Kaart T., Möls T. (2010). *Populatsioonigeneetika genotüüpide tasemel*. Loengukonspekt. Tartu Ülikool. [http://ph.emu.ee/~ktanel/MTMS\\_02\\_007/loeng\\_01\\_2010web.pdf](http://ph.emu.ee/~ktanel/MTMS_02_007/loeng_01_2010web.pdf) (09.05.2017).
- [9] Heinaru, A. (2012). *Geneetika*. Õpik kõrgkoolile. Tartu: Tartu Ülikooli Kirjastus. <http://geneetika.ee/geneetika-opik/> (09.05.2017).
- [10] Kasela S. (2011). *Ülegenoomne assotsiatsiooniuuring ja selle praktiline läbiviimine TÜ Eesti Geenivaramu andmete põhjal*. Bakalaureusetöö. Tartu Ülikool, matemaatilise statistika instituut.
- [11] Komar, A. A. (2009). *Single Nucleotide Polymorphisms: Methods and Protocols*. Second Edition. New York: Humana Press. <https://link.springer.com/book/10.1007%2F978-1-60327-411-1> (09.05.2017).
- [12] Dirk Van den Abele Ornitho-Genetics VZW. *Agapornis Genome Study*. [http://www.ornitho-genetics.info/?page\\_id=9334](http://www.ornitho-genetics.info/?page_id=9334) (09.05.2017).
- [13] Mändul M. (2016). *Geneetilise tagasiside mõju ravitulemusele*. Bakalaureusetöö. Tartu Ülikool, matemaatilise statistika instituut.
- [14] Tartu Ülikooli tervishoiu instituut. Rahvatervishoiu eesti-inglise seletav sõnastik. <http://rahvatervis.ut.ee/terms/terminid.php?id=784> (09.05.2017).

- [15] Käärik, E. (2013). *Andmeanalüüs II*. Loengukonspekt. Tartu Ülikool, matemaatilise statistika instituut. <http://dspace.ut.ee/bitstream/handle/10062/35401/AndmeanaluusII.pdf> (09.05.2017).
- [16] Möls, M. (2010). *Dispersioonanalüüsi segamudelid*. Loengukonspekt. Tartu Ülikool, matemaatilise statistika instituut. [http://www-1.ms.ut.ee/mart/sega2010/loeng2\\_3.pdf](http://www-1.ms.ut.ee/mart/sega2010/loeng2_3.pdf) (09.05.2017).
- [17] Wikipedia. *t-statistic*. <https://en.wikipedia.org/wiki/T-statistic> (09.05.2017).
- [18] dbSNP [e-andmebaas], <https://www.ncbi.nlm.nih.gov/snp/> (09.05.2017).
- [19] Exome Aggregation Consortium [e-andmebaas]. <http://exac.broadinstitute.org/> (09.05.2017).
- [20] Liina, T. (2014). *Mitmese testimise probleem*. Bakalaureusetöö. Tartu Ülikool, matemaatilise statistika instituut.
- [21] Kaart, T. (2009). *Statistiline andmetöötlus*. Loenguslaidid. [http://ph.emu.ee/~ktanel/VL\\_0435/magloeng3.pdf](http://ph.emu.ee/~ktanel/VL_0435/magloeng3.pdf) (09.05.2017).

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Pille-Riin Vare,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Depressiooniravi efektiivsus ja selle võimalik geneetiline taust: TÜ Eesti Geenivaramu poolt läbi viidud uuringu analüüs“, mille juhenda on Krista Fischer,
  - 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 09.05.2017